The role of pro regions in protein folding

David Baker, Andrew K Shiau and David A Agard

University of California, San Francisco, USA

In vivo, many proteases are synthesized as larger precursors. During the maturation process, the catalytically active protease domain is released from the larger polypeptide or pro-enzyme by one or more proteolytic processing steps. In several well studied cases, amino-terminal pro regions have been shown to play a fundamental role in the folding of the associated protease domains. The mechanism by which pro regions facilitate folding appears to be significantly different from that used by the molecular chaperones. Recent results suggest that the pro region assisted folding mechanism may be used by a wide variety of proteases, and perhaps even by non-proteases.

Current Opinion in Cell Biology 1993, 5:966–970

Introduction

An increasing number of proteases have been found to be synthesized as pro-enzymes, that is, the catalytically active protease region is eventually cleaved from a larger precursor polypeptide. Essentially all known extracellular bacterial proteases are made as pre-pro-proteins (the pre region being a signal sequence), and a growing number of intracellular and extracellular eukaryotic proteases have also been shown to be synthesized as pro-enzymes. Pro regions can be amino-terminal extensions, carboxyl-terminal extensions, or a combination of the two, but amino-terminal extensions are the most common. In addition, there is tremendous variability in the size of the pro regions, which range from ~40 amino acids to ~60 kDa.

In numerous cases, experiments have revealed that direct expression of the protease domain alone does not lead to the synthesis of active enzyme. As pro regions seem to be required transiently for the production of active proteases, it is reasonable to propose that the pro regions are involved in the maturation or folding of their associated protease domains. In several well characterized cases, it has been conclusively demonstrated that the pro regions are, indeed, required for the folding of their associated protease domains. Here we review the recent literature implicating pro regions in protein folding, and discuss possible mechanisms through which they may act.

Pro regions and folding

The two most extensively studied cases of pro region-assisted folding involve evolutionary unrelated prokaryotic serine proteases, α-lytic protease from the Gram negative soil bacterium *Lysobacter enzymogenes*, and subtilisin from the Gram positive bacterium *Bacillus subtilis*. α-Lytic protease has a 166-amino-acid amino-terminal pro region [1], and subtilisin has an amino-terminal pro region consisting of 77 residues [2]. In both cases, proper folding of the mature protease domains *in vivo* requires the corresponding pro region, supplied either in *cis*, as in the natural protease precursor [2,3], or in *trans* as part of a separate polypeptide chain [4]. These *in vivo* results have been reproduced *in vitro*; only in the presence of their pro regions do both proteases refold to their native states after denaturation [5•,6].

Several other proteases have pro regions that may well play important roles in folding. *In vitro*, pro-carboxypeptidase Y (pro-CPY) refolds readily following denaturation, while mature CPY alone refolds poorly [7]. *In vivo* data suggest roles for pro regions in the folding of the thermophilic protease aqualysin I [8•] and the *Yarrowia lipolytica* alkaline extracellular protease [9]. Expression experiments with the cysteine proteases papain [10], cruzain [11], and cathepsin L [12] also imply a requirement for pro regions.

Pro regions and protein transport

Although often small in size, pro regions can play a striking multitude of roles. For example, many of the pro regions mentioned above also function in intracellular protein transport. The pro regions of CPY [13] and several other yeast proteases are responsible for targeting the enzymes to the yeast vacuole. A striking recent result is that the *Yarrowia* protease pro region can function in *trans* to direct the protease through the secretory pathway [14••]. The connection between pro

Abbreviations

BPTI—bovine pancreatic trypsin inhibitor; **CPA**—carboxypeptidase A; **CPY**—carboxypeptidase Y; **NMR**—nuclear magnetic resonance.

 ISSN 0955-0674

regions and secretion is not restricted to eukaryotes: proper extracellular secretion of α-lytic protease requires a functional pro region [15•].

Pro regions as protease inhibitors

In all cases thus far examined, pro regions that promote folding have been found to be potent inhibitors of their respective enzymes. The α-lytic protease pro region inhibits mature α-lytic protease with a K_i of close to 10^{-11} M [5•]; the subtilisin pro region inhibits subtilisin with a K_i of 5×10^{-7} M [16]. Finally, pro-CPY has less than 0.1% of the enzymatic activity of mature CPY, suggesting that this pro region might also act as an inhibitor [7].

This correlation between folding activity and inhibitory activity suggests that the number of pro regions which function in folding may be considerably larger than presently thought. The pro regions of a large number of proteases have been shown to act as strong inhibitors, but their role in folding has not yet been examined. For example, the pro regions of the aspartyl proteases pepsin and cathepsin D inhibit the mature enzymes [17], the pro region of the metalloprotease carboxypeptidase A inhibits CPA with a K_i of 10^{-10} M [18], and the 62-amino-acid pro region of the cysteine protease cathepsin B inhibits cathepsin B action with a K_i of 4×10^{-10} M [19].

Evolution of pro regions

Several intriguing features concerning the evolution of pro regions may be inferred from Table 1, which gives a list of proteases having pro regions with known functions. Within the serine protease family, three evolutionary unrelated enzymes (α-lytic protease, subtilisin, and carboxypeptidase Y) have pro regions which function in folding. The fact that these serine proteases utilize a similar catalytic mechanism (based on a catalytic triad of active-site residues arranged in virtually identical relative positions on very different protein scaffolds) is the textbook example of convergent evolution. Perhaps an even more dramatic convergent evolution of folding mechanisms is suggested by the common requirement for pro regions. Pro region-dependent folding may not be restricted to the serine proteases: members of the three other major protease families also have pro regions that function as inhibitors (Table 1), and, by analogy, perhaps function in folding as well.

Interestingly, there are exceptions to this functional convergence, even within the serine protease family. The closest eukaryotic relatives of α-lytic protease, chymotrypsin, elastase and their homologues are produced via the classical zymogen activation mechanism [20]. These proteases can be expressed without the short polypeptide segments associated with zymogen activation, and hence do not require them for proper folding [21]. Thus, remarkably, the members of the α-lytic protease/chymotrypsin family, proteins with fairly similar sequences and nearly identical folds, possess evolutionarily divergent folding mechanisms.

Table 1. Proteases with pro regions for which a function is known.

Protease	N-terminal pro region?	Inhibitory activity?	Required for folding?
Serine protease			
α-lytic protease	+	+	+
Trypsin	(zymogen	–	–
Elastase	activation)	–	–
Chymotrypsin	"	–	–
Subtilisin	+	+	+
Aqualysin	+	nd	+
Carboxypeptidase Y	+	+	+
Y. lipolytica alkaline protease	+	nd	+
Metalloproteases			
Carboxypeptidase A	+	+	nd
Cysteine proteases			
Cathepsin B	+	+	nd
Cathepsin L	+	nd	+•
Cruzain	+	nd	+•
Papain	+	nd	+•
Aspartyl proteases			
Pepsin	+	+	nd
Cathepsin D	+	+	nd

+, present; –, absent; nd, not determined.
•Deletion of the pro region blocks production of active proteases *in vivo*. This could be a consequence of effects on folding and/or secretion

Mechanism of action

Recent work has shed some light on the mechanism by which pro regions facilitate folding. Protein folding in general involves a kinetic competition between 'on' pathway reactions leading to the folded state, and 'off' pathway reactions such as aggregation. Pro regions could facilitate folding either by increasing the rate of the forward folding reaction or by decreasing the rate of aggregation reactions. The molecular chaperones, which include such proteins as heat shock protein (hsp)70 and groEL/groES, are thought to function primarily by preventing aggregation [22,23]. Several lines of evidence suggest that, in contrast to the chaperones, pro regions function by directly increasing the rate of the forward folding reaction. First, as discussed above,

pro regions generally interact strongly with the product of the folding reaction, the native state (K_is ranging from 10^{-11} to 10^{-7} M). A second, stronger piece of evidence is that pro regions are required for folding under conditions in which off pathway reactions are almost completely suppressed. Denatured α-lytic protease cannot refold to the native state in the absence of the pro region [5•]. Instead, upon removal of denaturant, the protein folds to an intermediate state that has substantial secondary structure but little organized tertiary structure [24••]. The folding intermediate is stable for months, with no detectable conversion to the native state. Upon addition of the pro region, however, the intermediate is rapidly converted to the native state. As folding competence is maintained for an extended period of time, off pathway reactions have clearly been minimized under these conditions. Furthermore, this intermediate appears to be kinetically trapped, and the pro region appears to function by lowering the very high free-energy barrier which blocks access to the native state. Very similar observations have recently been reported for subtilisin [25•,26•], suggesting a similar role in folding for the subtilisin pro region.

What is the nature of the barrier blocking folding in the absence of the pro region? At least three quite different possibilities present themselves. First, the transition state in the folding reaction may correspond to a high-energy, distorted form of the native state, as suggested for other protein folding reactions. The observation that pro regions bind tightly to protease native states is most consistent with this possibility, since similar interactions could well be used in both stabilizing a native-like transition state and in binding to the native state. A late energy barrier is also suggested in the case of α-lytic protease by the observation that the folding intermediate that requires the pro region possesses much of the secondary structure of the native state [24••]. Second, the rate-limiting step may involve breaking inappropriate non-native interactions in a folding intermediate, in effect, an unfolding step. In wild-type subtilisin [25•] a highly structured intermediate accumulates in the absence of the pro region, which is only slowly converted to the native state (halftime ~8 days) upon addition of the pro region. In contrast, for a mutant lacking the high-affinity calcium-binding site [26•,27], a structureless intermediate accumulates in the absence of the pro region. The unstructured intermediate can then be rapidly folded by adding the pro region (folding halftime ~one minute). Taken together, the data indicate that destabilization of a folding intermediate can reduce the barrier to folding, and result in an increased reaction rate. Third, the barrier may be entirely entropic. In the absence of the pro region, the polypeptide chain may have access to a vast number of conformations of roughly equal energy, and thus be unable to find the native state because of the size of the search space. The pro region in this scenario would provide a path through this vast conformational space. This might well be the case for the mutant subtilisin lacking the calcium-binding site [26•]. The folding intermediate captured in the absence of the pro region possesses little secondary structure, and the slow step in the folding reaction catalyzed by the pro region appears to be the formation of the bimolecular complex.

What is the chemical nature of the interactions in the transition state? Some evidence exists, for both subtilisin [27••] and carboxypeptidase Y [7], that electrostatic interactions make a significant contribution to the energy barrier. While neither denatured subtilisin nor denatured CPY refold in low ionic strength buffer, some activity is regained in both cases when the proteins are allowed to refold in high ionic strength buffer. The rate-limiting step in these two cases may involve bringing together charged side chains.

Why pro regions?

Why should so many proteases have evolved a folding mechanism dependent on pro regions? The answer may involve the fact that proteases must function in very hostile environments (surrounded by other proteases), and thus need to be exceptionally stable. This requirement demands a large energy barrier to unfolding which, in the absence of pro regions, would also serve to block the forward folding reaction. Pro regions provide a means to lower transiently the height of the barrier during folding, and if, as in the case of the α-lytic protease pro region, they are designed to be exquisitely protease-sensitive, they will be destroyed once folding is complete. Their destruction thus prevents catalysis of the reverse reaction.

Future directions

Because pro regions function by reducing the height of kinetic barriers, and not by altering folding reaction equilibria (the native states of proteases are quite stable in the absence of their pro regions), they may be viewed as enzymes which catalyze the folding of their associated mature domains. As in the case of more traditional enzyme-mediated reactions, understanding the mechanism of catalysis requires understanding, firstly, the structure of the reaction substrate(s), i.e. the folding intermediate(s) which interact with the pro regions, secondly, the interactions between enzyme and substrate(s), i.e. the interactions between the pro region and the folding intermediate(s), and thirdly, the rate-limiting transition state and how the pro region lowers its energy. Answering these questions will require a combination of molecular genetics, kinetic analyses and, most importantly, structural analyses through nuclear magnetic resonance (NMR) and X-ray crystallography. Initial experiments along these lines have already begun. Polymerase chain reaction mutagenesis has been used in combination with a genetic screen to identify mutations which abolish the folding activity of the subtilisin pro region, as well as to identify a second site suppressor within the mature domain to a partic-

ular deleterious mutation [28•]. Investigations into the kinetics of folding assisted by pro regions and the energetics of binding (through calorimetry) have also taken place in the case of subtilisin [26•]. Finally, studies are in progress on the α-lytic protease folding reaction employing NMR techniques (JL Sohl and DA Agard, unpublished results).

Non-protease pro regions

Proteases are not the only proteins to be synthesized with large amino-terminal pro regions. For example, the growth hormones transforming growth factor-β1 (TGF-β1) and activin A are synthesized as parts of precursors containing amino-terminal pro regions. *In vivo* expression studies have shown that the pro regions are required for the folding and secretion of both hormones [29,30]. As in the case of the serine proteases described above, the pro regions can function either in *cis* or in *trans* (expressed using a separate promoter) [29]. A further similarity is that the TGF-β1 pro region associates with the mature growth factor and inhibits its activities [31]. It will be interesting to see how far the similarities continue.

Non-covalent interactions are not the only means by which pro regions can influence the folding of their associated mature domains. One of the most intensively studied protein folding pathways, that of bovine pancreatic trypsin inhibitor (BPTI), provides a striking example [32••]. The rate of folding to the final, three disulfide-bonded species is determined by disulfide chemistry. Recently, it was shown that the single cysteine of the BPTI pro region significantly increases the rate of folding by serving as an intramolecular disulfide reagent and promoting the formation of the proper disulfide bonds.

Conclusions

The surprising finding that pro regions function in many cases to facilitate the folding of their associated mature domains provides an exciting new twist on the conclusion from Anfinsen's classic experiments with ribonuclease, that the amino acid sequence of a protein determines its three-dimensional structure [33]. Although there exists an exceptionally stable conformation for the mature amino acid sequence (the native state), the mature protein may not be able to fold to this state on experimental time scales (as in the case of α-lytic protease). Thus, thermodynamic stability does not guarantee kinetic accessibility. Instead, nature has employed an additional stretch of polypeptide chain, the pro region, to solve the kinetic problem and to provide access to the native state.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. SILEN JL, MCGRATH CN, SMITH KR, AGARD DA: **Molecular Analysis of the Gene Encoding α-Lytic Protease: Evidence for a Preproenzyme.** *Gene* 1988, 69:237–244.
2. IKEMURA H, TAKAGI H, INOUYE M: **Requirement of Pro-Sequence for the Production of Active Subtilisin E in *Escherichia coli*.** *J Biol Chem* 1987, 262:7859–7864.
3. SILEN JL, FRANK D, FUJISHIGE A, BONE R, AGARD DA: **Analysis of Prepro-α-Lytic Protease Expression in *Escherichia coli* Reveals that the Pro Region is Required for Activity.** *J Bacteriol* 1989, 171:1320–1325.
4. SILEN JL, AGARD DA: **The α-Lytic Protease Pro-Region Does not Require a Physical Linkage to Activate the Protease Domain *in Vivo*.** *Nature* 1989, 341:462–464.
5. • BAKER D, SILEN JL, AGARD DA: **Protease Pro Region Required for Folding is a Potent Inhibitor of the Mature Enzyme.** *Proteins* 1992, 12:339–344.

Pro-region dependent folding of α-Lytic protease is reconstituted *in vitro*, and the pro region is found to be a potent inhibitor of the mature enzyme.

6. ZHU XL, OHTA Y, JORDAN F, INOUYE M: **Pro-Sequence of Subtilisin Can Guide the Refolding of Denatured Subtilisin in an Intermolecular Process.** *Nature* 1989, 339:483–484.
7. WINTHER JR, SORENSEN P: **Propeptide of Carboxypeptidase Y Provides a Chaperone-Like Function as Well as Inhibition of the Enzymatic Activity.** *Proc Natl Acad Sci USA* 1991, 88:9330–9334.
8. • LEE YC, OHTA T, MATSUZAWA H: **A Non-Covalent NH2-Terminal Pro-Region Aids the Production of Active Aqualysin I (a Thermophilic Protease) without the COOH-Terminal Pro-Sequence in *Escherichia coli*.** *FEMS Microbiol Lett* 1992, 91:73–77.

The precursor of aqualysin I, an extracellular protease produced by *Thermus aquaticus*, contains amino-terminal and carboxyl-terminal pro-sequences. The amino-terminal pro sequence is required either in *cis* or in *trans* for the production of active enzyme *in vivo*. The carboxyl-terminal pro region is not required for folding and actually inhibits rescue of folding by the amino-terminal pro region in *trans*.

9. FABRE E, NICAUD JM, LOPEZ MC, GAILLARDIN C: **Role of the Proregion in the Production and Secretion of the *Yarrowia lipolytica* Alkaline Extracellular Protease.** *J Biol Chem* 1991, 266:3782–3790.
10. VERNET T, KHOURI HE, LAFLAMME P, TESSIER DC, MUSIL R, SALIN BJ GOUR, STORER AC, THOMAS DY: **Processing of the Papain Precursor. Purification of the Zymogen and Characterization of Its Mechanism of Processing.** *J Biol Chem* 1991, 266:21451–21457.
11. EAKIN AE, MCGRATH ME, MCKERROW JH, FLETTERICK RJ, CRAIK CS: **Production of Crystallizable Cruzain, the Major Cysteine Protease from *Trypanosoma cruzi*.** *J Biol Chem* 1993, 268:6115–6118.
12. SMITH SM, GOTTESMAN MM: **Activity and Deletion Analysis of Recombinant Human Cathepsin L Expressed in *Escherichia coli*.** *J Biol Chem* 1989, 264:20487–20495.
13. VALLS LA, WINTHER JR, STEVENS TH: **Yeast Carboxypeptidase Y Vacuolar Targeting Signal is Defined by Four Propeptide Amino Acids.** *J Cell Biol* 1990, 111:361–368.
14. •• FABRE E, THARAUD C, GAILLARDIN C: **Intracellular Transit of a Yeast Protease is Rescued by *trans*-Complementation with its Prodomain.** *J Biol Chem* 1992, 267:15049–15055.

Reference [9] demonstrates that deletions within the pro region of the alkaline extracellular protease precursor of the yeast *Yarrowia lipolytica* interfere with the intracellular transit of the protease. This paper demonstrates that, rather remarkably, the proper secretion of the protease is restored when the pro domain is supplied in *trans*.

15. • FUJISHIGE A, SMITH KR, SILEN JL, AGARD DA: Correct Folding of α-Lytic Protease is Required for Its Extracellular Secretion from *Escherichia coli*. *J Cell Biol* 1992, 118:33–42.
Secretion of α-lytic protease from *E. coli* is shown to require proper folding of the protease by the pro region.

16. OHTA Y, HOJO H, AIMOTO S, KOBAYASHI T, ZHU X, JORDAN F, INOUYE M: Pro-Peptide as an Intramolecular Chaperone: Renaturation of Denatured Subtilisin E with a Synthetic Pro-Peptide. *Mol Microbiol* 1991, 5:1507–1510.

17. FUSEK M, MARES M, VAGNER J, VOBURKA Z, BAUDYS M: Inhibition of Aspartic Proteinases by Propart Peptides of Human Procathepsin D and Chicken Pepsinogen. *FEBS Lett* 1991, 287:160–162.

18. SAN SEGUNDO BS, MARTINEZ MC, VILANOVA M, CUCHILLO CM, AVILES FX: The Severed Activation Segment of Porcine Pancreatic Procarboxypeptidase A is a Powerful Inhibitor of the Active Enzyme. Isolation and Characterisation of the Activation Peptide. *Biochim Biophys Acta* 1982, 707:74–80.

19. FOX T, DE MIGUEL E, MORT JS, STORER AC: Potent Slow-Binding Inhibition of Cathepsin B by Its Propeptide. *Biochemistry* 1992, 31:12571–12576.

20. STROUD RM, KOSSIAKOFF AA, CHAMBERS JL: Mechanisms of Zymogen Activation. *Annu Rev Biophys Bioeng* 1977, 6:177–193.

21. HIGAKI JN, EVNIN LB, CRAIK CS: Introduction of a Cysteine Protease Active Site into Trypsin. *Biochemistry* 1989, 28:9256–9263.

22. CRAIG EA: Chaperones: Helpers Along the Pathways to Protein Folding. *Science* 1993, 260:1902–1903.

23. AGARD DA: To Fold or Not to Fold. *Science* 1993, 260:1903–1904.

24. •• BAKER D, SOHL JL, AGARD DA: A Protein-Folding Reaction under Kinetic Control. *Nature* 1992, 356:263–265.
A putative folding intermediate is trapped by omitting the pro region in an α-lytic protease refolding reaction. The intermediate, which is expanded in radius but has substantial native-like secondary structure, is stable for weeks at physiological pH, but rapidly folds to the active, native state on addition of the pro region. Because both the intermediate and native states are stable under identical conditions with no detectable interconversion, the folding of α-lytic protease must be under kinetic, and not thermodynamic, control.

25. • EDER J, RHEINNECKER M, FERSHT AR: Folding of Subtilisin BPN′: Characterization of a Folding Intermediate. *Biochemistry* 1993, 32:18–26.
Attempts to refold subtilisin in the absence of its pro region lead to the isolation of a metastable species which, like the α-lytic folding intermediate, possesses native-like secondary structure but little tertiary structure. The subtilisin pro region triggers the folding of this intermediate to the native state.

26. • STRAUSBERG S, ALEXANDER P, WANG L, SCHWARZ F, BRYAN P: Catalysis of a Protein Folding Reaction: Thermodynamic and Kinetic Analysis of Subtilisin BPN Interactions with Its Propeptide Fragment. *Biochemistry* 1993, 32:8112–8119.
Study of the pro region mediated folding of subtilisin is facilitated by taking advantage of a calcium-free mutant of subtilisin which, in contrast to wild-type subtilisin, is readily refolded by the pro region. The results suggest that the pro region functions early in folding; neither the unfolded enzyme nor the pro region have significant secondary or tertiary structure in isolation, and the rate limiting step in the folding reaction appears to be the formation of the initial collision complex between the two proteins.

27. •• BRYAN P, ALEXANDER P, STRAUSBERG S, SCHWARZ F, LAN W, GILLILAND G, GALLAGHER DT: Energetics of Folding Subtilisin BPN′. *Biochemistry* 1992, 31:4937–4945.
The subtilisin high-affinity calcium A site is mutated to allow study of folding independent of calcium binding. Unlike wild-type subtilisin, the mutant folds in the absence of the pro region. The rate of folding increases dramatically with increasing ionic strength, suggesting that part of the energy barrier is due to electrostatic interactions.

28. • KOBAYASHI T, AND INOUYE M: Functional Analysis of the Intramolecular Chaperone. Mutational Hot Spots in the Subtilisin Pro-Peptide and a Second-Site Suppressor Mutation within the Subtilisin Molecule. *J Mol Biol* 1992, 226:931–933.
PCR mutagenesis was used to generate mutants in the subtilisin pro region which affect the level of active enzyme. These mutations were concentrated in the hydrophobic regions of the pro-peptide. A second site suppressor of one of the pro region mutations was identified within the mature protease region.

29. GRAY AM, MASON AJ: Requirement for Activin A and Transforming Growth Factor-β1 Pro-Regions in Homodimer Assembly. *Science* 1990, 247:1328–1330.

30. SHA X, YANG L, GENTRY LE: Identification and Analysis of Discrete Functional Domains in the Pro Region of Pre-Pro-Transforming Growth Factor β1. *J Cell Biol* 1991, 114:827–839.

31. GENTRY LE, NASH BW: The Pro Domain of Pre-Pro-Transforming Growth Factor β1 When Independently Expressed is a Functional Binding Protein for the Mature Growth Factor. *Biochem* 1990, 29:6851–6857.

32. •• WEISSMAN JS, KIM PS: The Pro Region of BPTI Facilitates Folding. *Cell* 1992, 71:841–851.
The amino-terminal pro region of bovine pancreatic trypsin inhibitor (BPTI) is demonstrated to direct the folding of the mature domain *in vitro*. Through its single cysteine residue, the pro region is shown to act as an intramolecular thiol-disulfide reagent. Interestingly, a carboxyl-terminal single cysteine, tethered by a series of Ser-Gly-Gly residues, can substitute for the pro region.

33. ANFINSEN CB: Principles that Govern the Folding of Protein Chains. *Science* 1973, 181:223–230.

D Baker, AK Shiau and DA Agard, Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448, USA.